RECONOCIMIENTO Y LOCALIZACIÓN DE OBJETOS EN ACTIVIDADES HUMANAS

Proyecto Final de Carrera de Ingeniería Industrial

Especialidad en Electrónica y Automática

Autor: Elvira Alcázar Morán

Director: Antonio Sanchéz Salmerón y Co-director: Basam Musleh Lancis

10 de octubre de 2012

Agradecimientos

Índice general

Índice de figuras

Parte I

Memoria

Capítulo 1

Introducción

1.1. Resumen

En el presente proyecto de fin de carrera de Ingeniería Industrial se trata una nueva manera de identificación de objetos en vídeos, basada en la interacción persona-objeto, y no en la forma de estos como tradicionalmente se ha hecho. Para ello, se dejan de lado las técnicas habituales de identificación de objetos, y se usa como herramienta principal un sistema experto construido con una red dinámica de bayes.

El proyecto toma como conjunto de datos a estudiar un total de 110 vídeos de situaciones comunes de oficina, como por ejemplo, personas usando ordenadores, hablando por teléfono o leyendo papeles. El objetivo del proyecto es poder llegar a distinguir si en el vídeo analizado hay una persona hablando por teléfono o no, para lo cual se ha usado una red dinámica bayesiana. La construcción de esta ha sido la parte más importante del proyecto, donde ha sido necesario el etiquetado de todos los vídeos para extraer de ahí los nodos, el grafo y las probabilidades condicionales y temporales.

Como herramienta principal se ha usado MatLab. Se ha utilizado para el etiquetado de

vídeos (registro de las acciones que ocurren en cada diapositiva), cálculo de probabilidades e inferencia de la red de bayes.

1.2. Motivación

Un parte interesante de este proyecto es que con él, aporto un pequeño grano de arena a la participación por parte del departamento de automática de la Universidad Politécnica de Valencia en el concurso HARL 2012 (http://liris.cnrs.fr/harl2012/). El objetivo de este es reconocer acciones humanas en vídeos que se destacan por una fuerte interacción entre persona y objeto. El concurso pide reconocer diez clases de acciones que son:

- una discusión entre una o más personas.
- una persona le da algo a otra.
- coger o dejar un objeto.
- una persona sale o entra de una oficina.
- una persona intenta entrar en una oficina pero no lo consigue.
- una persona abre una oficina y entra en ella.
- una persona deja una mochila abandonada.
- apretón de manos entre dos personas.
- una persona teclea un ordenador.
- una persona habla por teléfono.

Otra gran motivación es poder hacer algo diferente y que no está muy desarrollado. Lo más frecuente en la visión computacional es reconocer los objetos por su forma, color, etc...Pero si nos ponemos en el caso de reconocer objetos pequeños, como es un teléfono móvil, el propio ojo humano en muchas ocasiones no lo ve, pero sabe que está en la imagen por la secuencia de acciones que se han dado, es decir, se deduce que el teléfono está en la imagen porque se sabe que la persona está hablando por teléfono. Este proceso es el que se ha intentado reproducir con la red dinámica de bayes.

1.3. Entorno de trabajo

En este apartado se van a nombrar y explicar brevemente las herramientas que se han usado a lo largo de este proyecto.

1.3.1. MatLab

El programa con el que he desarrollado toda la parte práctica del proyecto es el conocido entorno de programación MATLAB. Es un programa para el desarrollo de algoritmos, el análisis de datos, la visualización y el cálculo numérico. Una de las razones por las que se ha escogido es por la simpleza de su lenguaje de programación, ya que este proyecto no requiere una programación avanzada pero si se necesitaba un lenguaje fácil y sencillo con el que manejar toda la información. Más información sobre este, se puede encontrar en [1].

1.3.2. LyX

Con esta herramienta se ha escrito la memoria del proyecto. Es un programa que permite la edición de texto usando LaTeX. Se ha elegido este editor de textos antes que los tradicionales por la simplicidad con la que se puede crear un documento largo y con formato riguroso ya que no hay que darle formato final. Más información se puede consultar en [2].

1.4. Estructura de la memoria

En este apartado se describe brevemente el contenido de los diferentes capítulos que componen la memoria:

Capítulo 1: Introducción.

En este apartado se explican de forma introductoria los objetivos y características principales del proyecto, así como una explicación detallada de las herramientas usadas y un breve resumen del contenido de la memoria.

Capítulo 2: Estado del arte.

En este capítulo, se hace un pequeño recorrido sobre la evolución de la visión por computador y en concreto sobre el reconocimiento de objetos en imágenes. También se recorre la evolución del reconocimiento de objetos por su uso y no por su forma.

Capítulo 3: Redes de Bayes.

Este apartado explica toda la teoría necesaria para comprender las redes dinámicas de bayes, principio elemental sobre el que se basa este proyecto.

Capítulo 4: Metodología empleada.

En este capítulo se explica la construcción de la red dinámica de bayes, tanto la

obtención del grafo como la inferencia. La obtención de las probabilidades se explica con detalle en el Apéndice B.

Capítulo 5: Experimentación.

Se presentarán los resultados obtenidos de forma que se permita evaluar el método desarrollado haciendo hincapié en las conclusiones más destacas de cada análisis.

Capítulo 6: Conclusiones.

Este apartado presenta el estudio final del desarrollo del proyecto, donde se exponen las conclusiones y objetivos obtenidos. También se realiza el estudio de posibles líneas de futuros trabajos.

Capítulo 2

Estado del arte

A continuación se lleva a cabo un breve recorrido por la historia de la visión por computador, para profundizar posteriormente en el análisis y procesamiento de imágenes y en el reconocimiento de objetos tanto por su forma y color como por su función, siendo esto último el objeto del proyecto. También se habla sobre el reconocimiento de actividades por la importancia que tiene en este proyecto y sobre la librería de visión OpenCV por el uso cada más extendido de esta en la visión por computación.

2.1. Teoría de Imágenes

El trabajo con imágenes es una parte fundamental dentro de los proyectos que trabajan en el reconocimiento de objetos en imágenes, ya que mediante el procesamiento de las mismas se logra la detección de los eventos predeterminados. Es por ello que en este apartado se explicarán las características más importantes que poseen las imágenes.

2.1.1. Definición de píxel

El píxel es la unidad más pequeña en la que se puede descomponer una imagen digital, ya sea una fotografía o un fotograma de vídeo. Los píxeles aparecen como pequeños cuadrados en color. Las imágenes se forman como una matriz rectangular de píxeles, donde cada píxel forma un punto diminuto en la imagen total. Es importante hablar sobre los modelos de color de un píxel, el más conocido es el RGB (Red-Green-Blue). Este modelo permite crear un color componiendo tres colores básicos: el rojo, el verde y el azul. En el modelo RGB es frecuente que se use un byte para representar la proporción de cada una de las tres componentes primarias. Así, de una manera estándar, la intensidad de cada una de las componentes se mide según una escala entre 0 y 255. De esta forma, cuando una de las componentes vale 0, significa que esta no interviene en la mezcla y cuando vale 255 significa que interviene aportando el máximo de ese tono. Por lo tanto, el rojo se obtiene con (255, 0, 0), el verde con (0, 255, 0) y el azul con (0,0, 255), obteniendo un color resultante monocromático. La combinación de dos colores en nivel 255 con un tercero en nivel 0 da lugar a tres colores intermedios: el amarillo (255, 255 ,0), el cian (0, 255, 255) y el magenta (255, 0, 255).

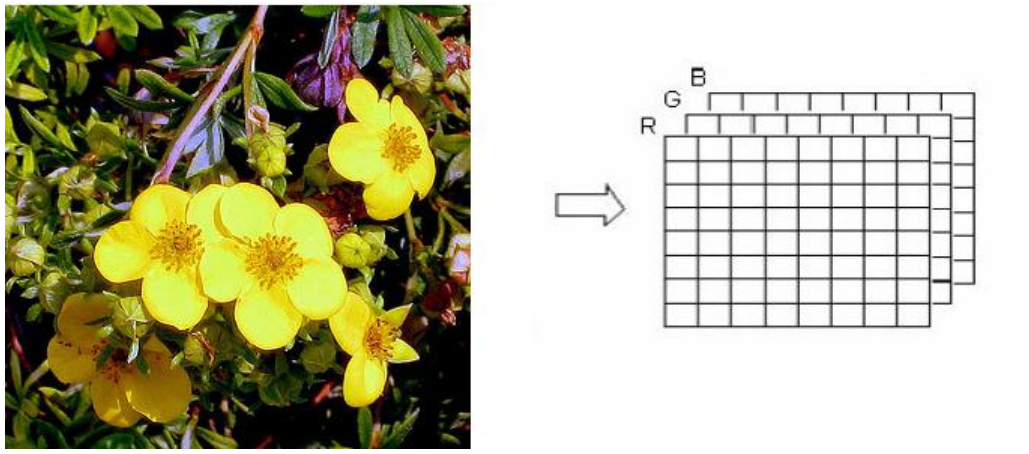

Figura 2.1: Composición de una imagen en RGB

2.1.2. Resolución de la imagen

La resolución de imágenes describe cuánto detalle puede observarse en una imagen. El término es comúnmente utilizado en relación a imágenes de fotografía digital, pero también se utiliza para describir la nitidez de la misma. Tener mayor resolución se traduce en obtener una imagen con más detalle o calidad visual y por tanto mayor información. Para las imágenes digitales almacenadas como mapa de bits, la convención es describir la resolución de la imagen con dos números enteros, donde el primero es la cantidad de columnas de píxeles y el segundo es la cantidad de filas de píxeles. La convención que le sigue en popularidad es describir el número total de píxeles en la imagen (usualmente expresado como la cantidad de megapíxeles), que puede ser calculado multiplicando la cantidad de columnas de píxeles por la cantidad de filas de píxeles.

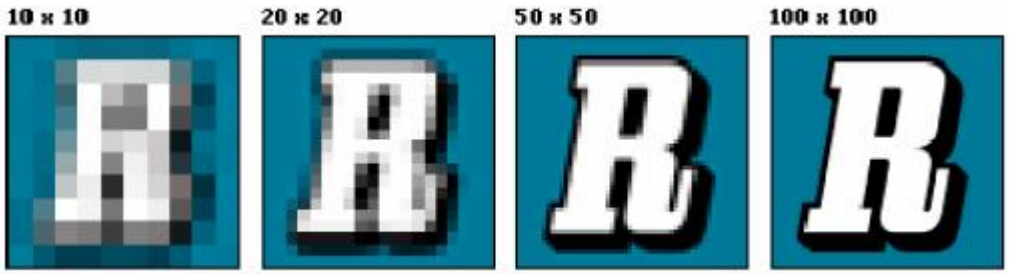

Figura 2.2: Distintas resoluciones en una misma imagen

2.2. Visión por computador

Desde los inicios de la informática, se ha perseguido la imitación del comportamiento humano en las máquinas. El cuerpo humano ha sido y es considerado como el modelo más eficiente a seguir para determinadas disciplinas. Hoy por hoy, comportamientos triviales para el hombre no están todavía científicamente tipificados, y entre ellos se encuentra el proceso de la visión humana. Por esta razón, la construcción de un sistema que emule el sistema visual humano sería prácticamente imposible. Todo esto, originó el abandono paulatino de esta línea de investigación, hasta que en la década de los ochenta se produce un replanteamiento en la especificación de los objetivos para esta disciplina. Se cambia la ambiciosa denominación de "Visión Artificial", por una más humilde y consecuente con el objetivo perseguido, como es "Visión por Computador". A este cambio en el punto de vista sobre esta disciplina se le unen, la existencia de computadores más potentes, elementos hardware específicos que relevan el uso de algoritmos complejos, y una mayor experiencia histórica en otros campos de la informática. Empiezan a describirse metodologías que dividen el problema de la visión por computador en distintas fases, cuya solución resulta más asequible, y lo relacionan con otras disciplinas, que hasta el momento eran independientes como pueden ser el procesamiento de imágenes, o el reconocimiento de patrones. La visión por computador actualmente comprende tanto la obtención como la caracterización e interpretación de las imágenes. Esto supone algoritmos de muy diversos tipos y complejidades. En un sistema de visión por computador actual se pueden distinguir seis etapas o fases:

1. Adquisición de la imagen.

2. Procesamiento. Incluye técnicas de reducción del ruido y realce de detalles.

3. Extracción de características.

4. Segmentación. Proceso que divide la imagen en objetos.

5. Descripción de objetos. Asocia un significado a un conjunto de objetos reconocidos.

6. Reconocimiento o clasificación.

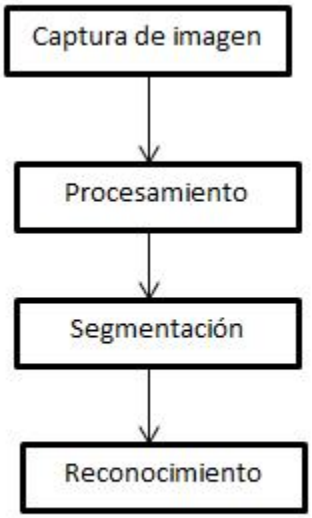

Figura 2.3: Esquema sistema de visión computacional

Esto supone distintos tipos de procesamiento en función del nivel en el que nos movamos:

- Visión de bajo nivel: comprende la captación y el pre-procesamiento. Ejecuta algoritmos típicamente de filtrado, restauración de la imagen, realce, extracción de contornos, etc.

- Visión de nivel intermedio: comprende la segmentación, descripción y reconocimiento, con algoritmos típicamente de extracción de características, reconocimiento de forma y etiquetado de éstas.

- Visión de alto nivel: comprende la fase de interpretación, normalmente estos algoritmos se refieren a la interpretación de los datos generalmente mediante procedimientos típicos de la Inteligencia Artificial para acceso a bases de datos, búsquedas, razonamientos aproximados, etc.

Por último cabe destacar el rango de aplicaciones en la que la visión por computador tiene cada vez más, un papel importante:

- Militares: gran parte de los logros informáticos conseguidos han sido promovidos o posteriormente adquiridos y mejorados por este sector. Entre las aplicaciones más comunes podemos destacar la detección y seguimiento de objetivos.

- Robótica: en aplicaciones industriales para guiado de robots.

- Análisis de imágenes tomadas por satélite.

- Identificación automática de huellas: muy extendidos en sistemas de seguridad y control de acceso.

- Control de calidad: muy extendido en cadenas de montaje.

- Medicina: dando especial importancia a los avances obtenidos con imágenes RMI (imágenes por resonancia magnética), y a los ya conocidos logros en imágenes por rayos X.

- Reconocimiento de Caracteres: dentro de esta área se encontrarían aplicaciones como lectura de etiquetas, procesamiento de cheques bancarios, lectura de texto, etc...

2.3. Reconocimiento de objetos

Al conjunto de técnicas que tienen por objetivo detectar objetos en movimiento se denomina detección de primer plano. Actualmente, el uso de sistemas inteligentes de análisis de secuencias de video está cada vez más presente en una gran variedad de sistemas como los de vídeo vigilancia [3]. Debido a la creciente cantidad de información visual generada por las cámaras y sensores de estos sistemas, es necesario desarrollar herramientas de análisis automático, que operen en tiempo real y que permitan extraer las regiones de interés de la secuencia de video analizada. En este contexto, el primer problema es la localización de la región donde sucede algo relevante. Esta operación suele conocerse como segmentación. El objetivo de la segmentación de objetos habitualmente consiste en diferenciar los objetos en movimiento del primer plano de una imagen, del resto de los objetos o fondo ("background"). En el caso de una escena grabada por una cámara fija, las técnicas de segmentación más eficaces son las basadas en el modelado y posterior sustracción del fondo. La segmentación automática de objetos presenta múltiples complicaciones, siendo una de las tareas más complicadas dentro del procesado de video.

2.3.1. Clustering Color

En un comienzo, el proyecto se intento orientar por este aspecto, pero al final se vio que era mucho más adecuado usar las redes dinámicas de bayes para nuestro propósito.

Es un tema importante dentro del reconocimiento de objetos así que merece la pena comentarlo en este capítulo.

El clustering color tiene como objetivo hacer la segmentación de objetos descrita en el apartado anterior basándose en los colores de la imagen. El algoritmo más conocido para ello, es el k-means. Este algoritmo se encuentra en la librería de visión OpenCV aunque hay que modificarlo para que actúe directamente sobre imágenes. Tal y como esta en la librería, se puede aplicar sobre un conjunto de datos. El k-means está dentro del conjunto de algoritmos llamados de agrupamiento, es una forma rápida y sencilla de dividir una base de datos en k grupos fijados con anterioridad. La idea principal es definir k centroides (uno para cada grupo) y luego tomar cada punto de la base de datos y situarlo en la clase de su centroide más cercano. El próximo paso es recalcular el centroide de cada grupo y volver a distribuir todos los objetos según el centroide más cercano. El proceso se repite hasta que ya no hay cambio en los grupos de un paso al siguiente. Un ejemplo de como actúa sobre un conjunto de datos en la figura 2.4.

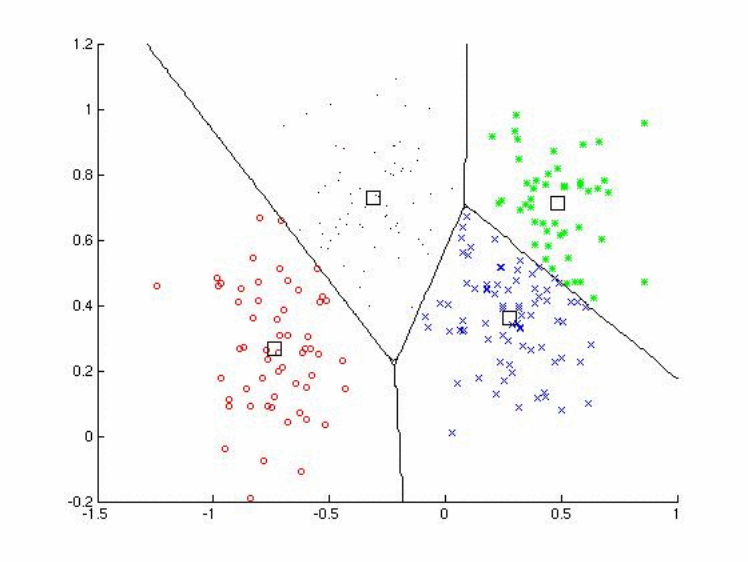

Figura 2.4: K-means sobre conjunto de datos

Esto aplicado a las imágenes es separar los objetos de la imagen por colores, de tal forma que cada grupo es un color, o mejor dicho, una gama de colores. Un ejemplo de como actúa el algoritmo sobre una imagen se puede observar en la figura 2.5, la primera foto es de la imagen antes de aplicarle el algoritmo y las sucesivas representa cada grupo de color en el que se ha dividido.

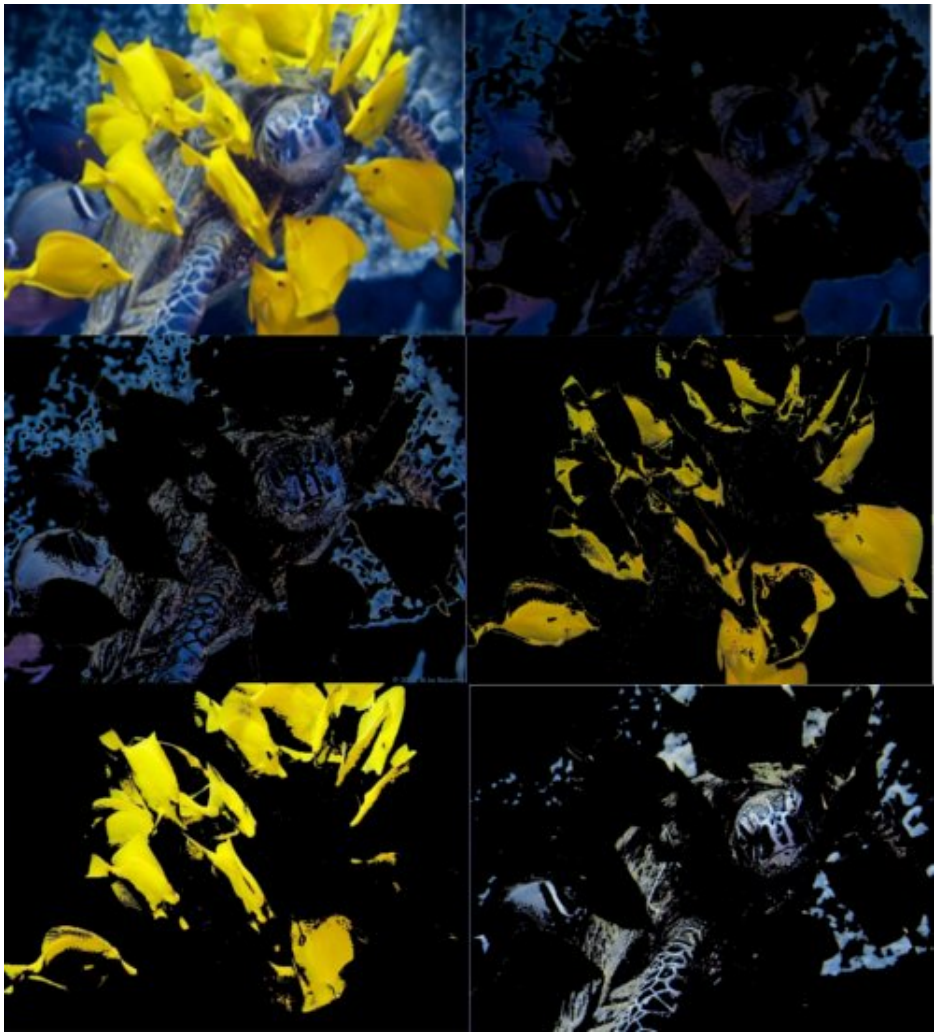

Figura 2.5: K-means sobre imagen

El mayor problema de este método es elegir los k grupos de antemano. Si se escogen pocos, no serán suficientes para hacer una segmentación por colores, mientras que si se

eligen muchos, tendremos problemas, pues dentro de un mismo objeto lo normal es que hayan varios colores y entonces estaremos segmentando el propio objeto a reconocer.

2.3.2. Reconocimiento de objetos por su uso

Este es el apartado más importante pues es el objeto de este proyecto. El documento clave que se estudio para llegar a la idea de la construcción de la red dinámica de bayes es [4]. En este, se presenta la idea de llegar al reconocimiento de los objetos a través de tres redes dinámicas de bayes. La función de cada red es la siguiente:

1. Estimación de la pose humana.
2. Reconocimiento de las acciones que interactúan con objetos.
3. Identificación del objeto.

La tercera red toma como datos los resultados obtenidos en la primera y en la segunda para su construcción e inferencia. Eso mismo es lo que se ha querido reproducir en este proyecto pero con más sencillez, pues los datos de las primeras redes de bayes se dan por conocidos y no hacen falta hallarlos. Otros aspectos que se recogen de este trabajo son:

- la separación en regiones según los objetos. Es decir, en cada imagen se etiquetarán los objetos por separado, habrá una red dinámica de bayes para cada objeto.
- la poca eficacia del método si lo que se quiere es reconocer varios objetos que tengas funciones similares.

2.4. Reconocimiento de acciones

Este es un apartado importante pues en este proyecto se da por supuesto que se saben las acciones que ocurren en cada momento en los vídeos. Lo ideal hubiese sido incluir el reconocimiento de acciones en este proyecto pero no era posible pues lo hacia demasiado extenso.

El reconocimiento de acciones humanas es un área de investigación muy activa. Existen numerosas técnicas que se aplican a esta tarea, bien basadas en la definición de un modelo del cuerpo humano o en características extraídas directamente de los vídeos; bien usando información sobre la forma del sujeto o sobre su patrón temporal (esta última sería la más importante para este proyecto ampliado). Los métodos propuestos para el reconocimiento de acciones se agrupan en dos categorías:

- Los que abordan la cuestión de forma tradicional, como un problema de emparejamiento de patrones. Se extrae un patrón de la secuencia de imágenes y se compara con un prototipo almacenado.

- Los modelos de espacio de estado, los cuales definen un espacio de estado en el cada punto de este, se representa un evento de la acción (pose, movimiento,...), así, se proporciona un modo compacto de describirla. Estos modelos pueden ser deterministas que conciben las acciones como procesos observables en el que el conjunto de estados es conocido; o probabilísticos, en los que a cada estado se le asocia una probabilidad, siendo la probabilidad conjunta del camino a través de los estados lo que conforma el modelo de cada acción.

Las técnicas primeramente mencionadas suelen tener un bajo coste computacional durante la clasificación, pero la extracción de características puede ser muy compleja, además,

son más susceptibles al ruido y a otros factores como el punto de vista de la cámara. En cambio, los modelos de espacio de estados, manejan de forma natural la variabilidad temporal en la ejecución de las acciones ya que el mismo estado puedo visitarse repetidamente. Entre los modelos de espacio de estados los que tienen mayor popularidad son los modelos ocultos de Markov (HMM) y sus variantes. Pues se ha probado con éxito para problemas como el reconocimiento de gestos, como en [5]. Los HMM se emplean para reconocer desde acciones sencillas como andar o correr hasta acciones complejas. En [6] se usan HMM jerárquicos para reconocimiento de acciones formadas por combinación de varias actividades, como por ejemplo, las tareas en una oficina.

2.5. OpenCV

OpenCV (Open source Computer Vision library) es una librería de código abierto para visión artificial desarrollada por Intel y publicada bajo licencia BSD, por lo que está permitido su uso tanto con fines de investigación como comerciales según las condiciones expresadas en ella en el año 2002. Es una librería multiplataforma, existiendo versiones compatibles con sistemas operativos Windows, Linux y Mac OS X y está optimizada para su uso en arquitecturas Intel con soporte para instrucciones MMX (a partir de Pentium II MMX). Además ha sido diseñada para ser usada junto con la librería de Intel IPL (Image Processing Library) para manejo de imágenes digitales a bajo nivel extendiendo su funcionalidad y aprovechando la optimización que implementa esta para arquitecturas Intel. OpenCV proporciona una gran variedad de abstracciones para su uso a alto nivel e implementa funciones y algoritmos para distintas técnicas de visión artificial. Entre sus funcionalidades cabe destacar procesado de imágenes, detección de rasgos, segmentación de objetos, calibración de cámara, análisis estructural, reconstrucción 3D y análisis de

movimiento. Además proporciona un sistema de manejo de errores, una interfaz gráfica y un sistema de adquisición de imágenes. Más información se puede encontrar en [8].

Capítulo 3

Redes Bayesianas

La base teórica del proyecto es una red dinámica bayesiana, para comprender este concepto es importante tener conocimiento de teoría de probabilidad básica, empezando por el teorema de Bayes.

3.1. Teorema de Bayes

El teorema de Bayes fue enunciado por Thomas Bayes en 1763 siendo este un resultado importante dentro de la teoría de la probabilidad. Este teorema expresa la probabilidad condicional de un evento aleatorio.

Se define como probabilidad condicional, la probabilidad de que ocurra un suceso A sabiendo que también sucede otro evento B. La manera formal de escribirlo es P(A|B) y la definición es:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Por el teorema de la multiplicación sabemos que si los sucesos A y B son independientes, entonces P(A∩B)=P(A)·P(B), por tanto P(A|B)=P(A) e igualmente ocurre con

P(B|A)=P(B). Un error muy común es asumir que P(A|B) y P(B|A) son casi iguales pero la verdadera relación entre estos dos términos se expresa en el teorema de Bayes con la siguiente ecuación:

$$P(A|B) = P(B|A) \cdot \frac{P(A)}{P(B)}$$

Este enunciado del teorema de Bayes es muy sencillo y sólo para dos sucesos, el teorema tal y como lo enuncio Thomas Bayes es:

Sea $\{A_1, A_2, ..., A_n\}$ *un conjunto de sucesos mutuamente excluyentes y exhaustivos, y tales que la probabilidad de cada uno de ellos es distinta de cero. Sea B un suceso cualquiera del que se conocen las probabilidades condicionadas P(A|B$_i$). Entonces, la probabilidad P(B$_i$|A) viene dada por la expresión:*

$$P(A_i \mid B) = \frac{P(B \mid A_i) \cdot P(A_i)}{P(B)}$$

donde:

- *P(A$_i$) son las probabilidades a priori.*
- *P(B|A$_i$) es la probabilidad de B en la hipótesis A$_i$.*
- *P(A$_i$|B) son las probabilidades a posteriori.*

Según el teorema de las probabilidades totales, la probabilidad de B se puede expresar como $P(B) = \sum_{i=1}^{n} P(B|A_i) \cdot P(A_i)$ y por tanto otra forma de enunciar el teorema de Bayes es:

$$P(A_i \mid B) = \frac{P(B \mid A_i) \cdot P(A_i)}{\sum_{i=1}^{n} P(B|A_i) \cdot P(A_i)}$$

Pongamos un ejemplo sencillo para comprender mejor esto. Supongamos una fábrica de tornillos en las que el 60 % son producidos por una máquina A y el 40 % por una máquina B. La proporción de defectuosos en A es del 0.1 y en B del 0.5. ¿Cuál sería la probabilidad de que un tornillo de dicha fábrica sea defectuoso? ¿Cuál es la probabilidad de que, sabiendo que un tornillo es defectuoso, proceda de la máquina A?.

Llamamos D al suceso tornillo defectuoso, así que como datos tenemos: p(A)=0.6, p(B)=0.4, p(D|A)=0.1 y p(D|B)=0.5. Lo primero que nos piden es la probabilidad de que un tornillo de la fábrica sea defectuoso, es decir: p(D). Aplicando el teorema de las probabilidades totales visto anteriormente obtenemos:

$$p(D) = p(D \mid A) \cdot p(A) + p(D \mid B) \cdot p(B) = 0{,}1 \cdot 0{,}6 + 0{,}5 \cdot 0{,}5 = 0{,}26$$

Lo segundo que nos piden es la probabilidad de que sabiendo que un tornillo es defectuoso, proceda de la máquina A, es decir, p(A|D). Para ello, aplicamos el teorema de Bayes:

$$p(A \mid D) = \frac{p(D \mid A) \cdot p(A)}{p(D)} = \frac{0{,}1 \cdot 0{,}6}{0{,}26} = \frac{3}{13}$$

Si pidiesen la probabilidad de que sabiendo que el tornillo es defectuoso procediera de la máquina B, sería:

$$p(B \mid D) = \frac{p(D \mid B) \cdot p(B)}{p(D)} = \frac{0{,}5 \cdot 0{,}5}{0{,}26} = \frac{25}{26}$$

Una vez explicado el teorema de Bayes, podemos explicar convenientemente las redes de Bayes, esto se desarrollará en el siguiente apartado.

3.2. Redes de bayes

3.2.1. Definición

Para el desarrollo del caso práctico del proyecto es necesario comprender la técnica de redes bayesianas. En el presente capítulo, se mostrarán aquellos aspectos más importantes que se han considerado necesarios para el entendimiento global de este tema. El formalismo de las redes bayesianas facilitará la representación eficiente y el razonamiento riguroso en situaciones en las que se dispone de conocimiento incierto. Actualmente, es una de las técnicas que dominan la investigación de la inteligencia artificial en el razonamiento incierto y en los sistemas expertos. Esta aproximación facilita el aprendizaje a partir de la experiencia, y combina lo mejor de la inteligencia artificial clásica y las redes neuronales.

Las redes bayesianas son una alternativa a la hora de implementar un sistema experto probabilístico ya que poseen ciertas cualidades que otras técnicas no permiten, como por ejemplo, admiten el aprendizaje sobre relaciones de dependencia y causalidad, permiten la combinación de conocimiento con datos, evitan el sobre-ajuste continuo de los datos y pueden manejar bases de datos incompletas.

Formalmente, una red bayesiana es un grafo acíclico dirigido (DAG) en el cual cada nodo representa una variable y cada arco una dependencia probabilística que especifica la probabilidad condicional de cada variable dados sus padres. La red bayesiana se puede ver como un conjunto formado por tres partes:

- un conjunto de variables del dominio que se quiere representar.
- un grafo acíclico dirigido (DAG) cuyos nodos están etiquetados con los elementos del anterior conjunto.
- una distribución conjunto sobre las variables.

Un ejemplo muy básico de red de bayes es la mostrada en la figura 3.1, ejemplo que determina si la hierba está mojada o no dependiendo de las variables aleatorias que describen el problema (en el ejemplo: nublado, aspersor, lluvia). Los nodos representan las variables del problema y los arcos las relaciones causales entre estas.

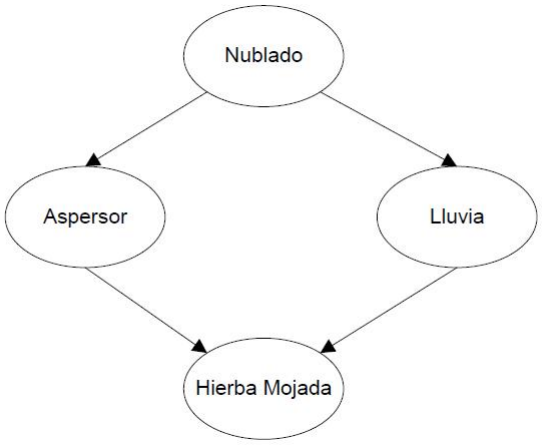

Figura 3.1: Diagrama simple de red de Bayes

El modelo más común son las redes de bayes discretas, que se caracterizan porque cada uno de sus variables sólo pueden tomar un conjunto determinado de valores. En la figura 3.2 se presenta un ejemplo de red bayesiana discreta, en el que además de presentar el diagrama de relaciones se muestran las distribuciones de probabilidad condicionada asociadas a los valores de las variables, donde cada variable tiene unos valores determinados. En este caso, los valores posibles de cada variable serán que se produzcan o no.

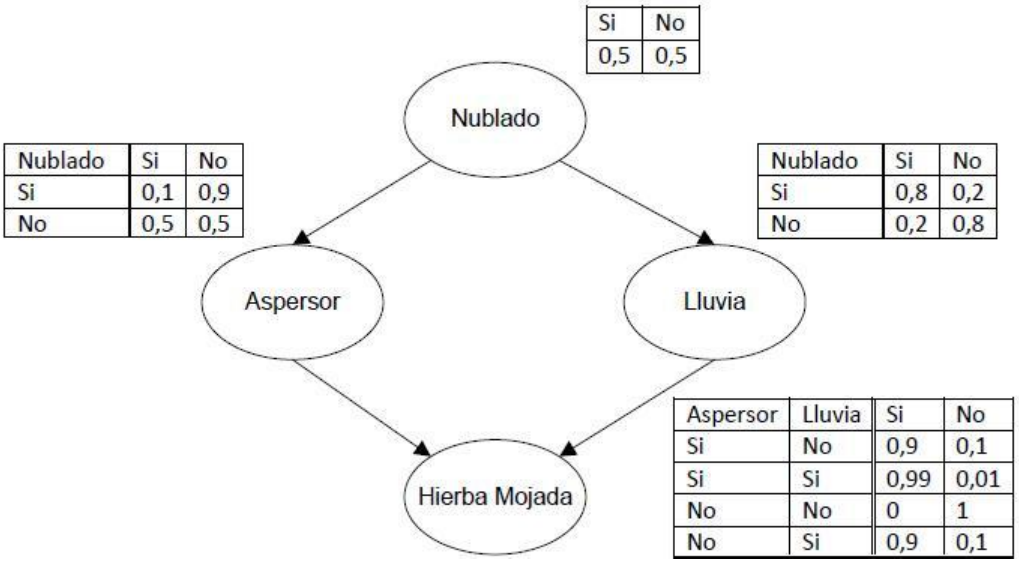

Figura 3.2: Red bayesiana discreta

3.2.2. Construcción

Para la construcción de una red bayesiana, es necesario realizar varias tareas hasta conseguir una estructura final dispuesta a funcionar dentro del sistema experto. Existen dos formas de construir una red bayesiana: de forma automática o de forma manual (también se puede la combinación de ambos tipos). Ambas formas, implican en su proceso de construcción básicamente tres tareas:

1. Identificar las variables y sus valores.

2. Identificar las relaciones entre las variables, completando la definición del grafo que representa el modelo.

3. Obtener las probabilidades asociadas a cada nodo del grafo.

El proceso manual, construye la red bayesiana a partir de la ayuda de un experto humano que conozca a fondo el problema que se quiere modelar. Así a través de esta ayuda, se establecerá primero la estructura de la red causal (nodos y estructura del grafo), y posteriormente añadirá las probabilidades condicionales de los nodos creados. El proceso automático consiste en tomar una base de datos en la que todas las variables que nos interesan estén representadas y que contenga un número de casos suficientemente grande. Entonces, aplicando ciertos algoritmos desarrollados especialmente para esta tarea, se obtienen los enlaces y las probabilidades condicionales que definen la red bayesiana. Sin embargo, en muchos problemas reales, es muy difícil contar con una base de datos suficientemente grande y detallada para la construcción de la red de esta forma.
En el caso específico de este proyecto la construcción de la red bayesiana será totalmente manual. Eligiendo nosotros los nodos, la estructura y calculando las probabilidades a partir del etiquetado de los vídeos, es por eso que a continuación se detallará de forma más extensa este tipo de construcción.

La construcción manual es complicada y no existen unos criterios definidos que se puedan aplicar siempre ante cualquier problema, siendo el sentido común y la experiencia previa sobre el tema dos factores fundamentales a la hora de la construcción, aún así, existen unos pasos generales que nos pueden ayudar a esta tarea. A grandes rasgos serían:

1. Obtener bibliografía sobre el dominio a trabajar.

2. Disponer de una herramienta de edición y procesado de redes bayesianas que recoja las necesidades del desarrollo.

3. Tener una base de datos amplia y completa que pueda usarse como apoyo para la construcción.

4. Definir el grafo, para lo cual debemos elegir los nodos y lo más importante, las relaciones entre ellos. No se debe olvidar que hay que definir los valores que puede tomar cada variable, lo más normal es que en una red de bayes discreta tomen dos valores como por ejemplo: Si-No, Aparece-No aparece, Presente-No presente...

5. Obtención de las probabilidades condicionadas de las variables que presenten nodos padres y de las probabilidades a priori para los nodos raíz. Esto es importante, por lo que se explicará con más detalle a continuación: si nos fijamos en la Figura 3.2, vemos que el nodo "Nublado" es el raíz y que las probabilidades a priori son las que aparecen a su lado, estas no dependen de nada, sólo de la cantidad de veces que el cielo este nublado o no. Los nodos "Aspersor" y "Lluvia" son nodos hijos de "Nublado" y a la vez padres del nodo "Hierba mojada", las probabilidades condicionadas de cada nodo son las que aparecen en la tabla que cada uno tiene a su lado, un ejemplo de estas probabilidades sería: p(Lluvia | Nublado), es decir, probabilidad de que sabiendo que está nublado haya lluvia, en nuestro ejemplo sería del 80 %.

En la figura 3.3 se muestra un esquema de estos pasos a seguir que se usará más adelante en el apartado práctico para la construcción de la red.

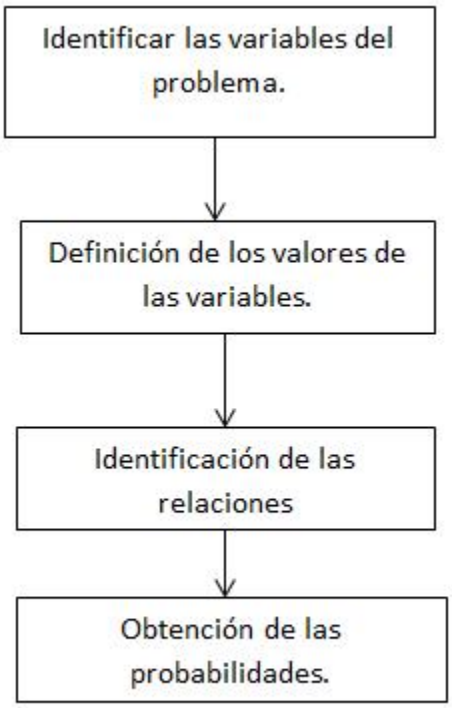

Figura 3.3: Esquema construcción red de bayes

3.2.3. Inferencia

Una vez construida la red (probabilidades incluidas), el siguiente paso es determinar el cambio producido en estas probabilidades cuando los valores de algunas de las variables llegan a ser conocidos, es decir, cuando introducimos evidencias (sabemos que una variable tiene un valor del 100 % o del 0 %, es decir, que ocurre o no ocurre). A este proceso se le denomina propagación de probabilidades y consiste en propagar los efectos de las evidencias a través de la red para conocer la probabilidad a posteriori de las variables. Existen diferentes tipos de algoritmos para el cálculo de las probabilidades posteriores

que no son necesarios explicar para el entendimiento de este proyecto.

3.3. Redes dinámicas de bayes

3.3.1. Definición

Para poder representar procesos dinámicos de modelos probabilísticos temporales, existe una extensión a las redes bayesianas explicadas anteriormente conocidas con el nombre de red bayesiana dinámica, RDB (en inglés "Dynamic Bayesian Network"). Estas redes se basan en la discretización del tiempo y en crear una réplica de cada variable aleatoria para cada punto temporal. La representación de las redes bayesianas dinámicas se basa en dos suposiciones importantes:

- **Suposición markoviana:** el futuro es condicionalmente independiente del pasado dado el presente. Esta propiedad no debe ser confundida con la condición de Markov que se aplica en la definición formal de una red bayesiana, aunque ambas estás muy relacionadas.
- **Proceso estacionario en el tiempo:** las probabilidades condicionales en el modelo no cambian con el tiempo.

Las dos suposiciones anteriores, implican que se pueda definir una RBD en base a dos componentes. El primero, como una red base estática que se repite en cada periodo, de acuerdo a cierto intervalo de tiempo predefinido y el segundo, en base a una red de transición entre etapas consecutivas.

A continuación se explica de forma breve una serie de consideraciones a tener en cuenta sobre los nodos, la estructura y las tablas de probabilidad condicionada que servirán

de ayuda para comprender el proceso de construcción e inferencia de una red dinámica de bayes.

Vamos a suponer que el dominio está compuesto por un conjunto de n variables aleatorias X={x$_1, x_2, ..., x_n$} siendo cada variable un nodo en la red, al construir una RBD temporal, el estado actual de la red sera representado por t, el estado anterior por t-1 y el estado próximo por t+1, de forma que cada nodo tendrá su representación en cada momento de tiempo. Hay que tener en cuenta, que en la mayoría de los casos, el valor de una variable en un tiempo t afecta a su valor siguiente en t+1 de manera circular, de tal forma que casi siempre este tipo de relaciones están presentes en la estructura de la red. Aunque en general, el valor de cualquier nodo puede afectar al valor de cualquier otro nodo en el instante de tiempo siguiente.

En la figura 3.4 se muestra un ejemplo simple de red dinámica de bayes, como se puede observar la estructura del árbol se repite en el tiempo.

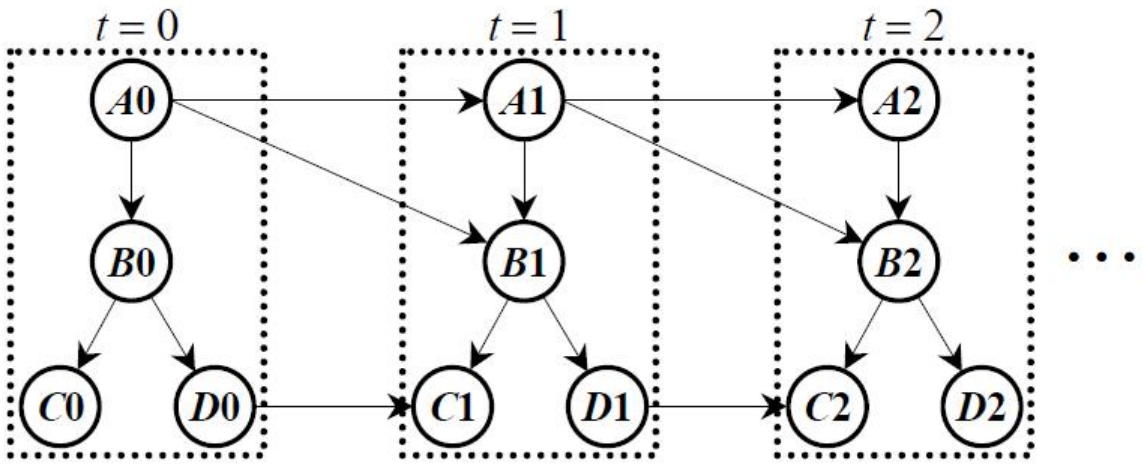

Figura 3.4: Red dinámica de bayes

3.3.2. Construcción

La construcción de las RDB es similar a el de las redes de bayes con la diferencia del aspecto temporal. Tomando como referencia en el esquema de la figura 3.3, las diferencias serán:

- A la hora de identificar las relaciones entre las variables será de especial interés definir correctamente las relaciones temporales entre variables.
- En el cálculo de probabilidades no sólo habrá que calcular las condicionales sino también las temporales.

El cálculo de las probabilidades temporales es algo nuevo por lo que vamos a explicarlo más detalladamente a continuación debido a su importancia. En el ejemplo de la figura 3.4, una probabilidad temporal que relaciona una misma variable sería: p(A1|A0), es decir, la probabilidad de que ocurra el evento A1 sabiendo que ha ocurrido el evento A0. Una probabilidad temporal relacionando distintas variables sería: p(C1|D0), es decir, la probabilidad de que ocurra C1 sabiendo que ha ocurrido el evento D0.

3.3.3. Inferencia

Como en el apartado anterior de construcción, ahora ocurre lo mismo, la inferencia en RDB es similar a la de las redes de bayes. Los métodos de inferencia para las probabilidades condicionales serán iguales mientras que la inferencia para las probabilidades temporales será lo nuevo a contemplar en este apartado.

Entonces, teniendo en cuenta las evidencias sobre un conjunto de nodos, $E_{1:t}$, desde el primer corte de tiempo, hasta un corte actual t, el proceso de actualización de las

evidencias en la RDB completa se puede realizar a través de algoritmos de inferencia estándares de las redes bayesianas. Esto significa que dada una secuencia de observaciones, uno puede construir la representación de la red bayesiana completa duplicando cortes hasta que la red sea lo suficientemente grande como para adecuarse a las observaciones, para posteriormente una vez completa la RDB usar cualquiera de los algoritmos de inferencia. Esta técnica se conoce como desenrollado. Sin embargo, cuando se tiene una secuencia elevada de observaciones , la técnica de desenrollado no es eficiente, y de este modo crecería sin límites conforme se añadieran más observaciones. Más aún, si cada vez que se añade una observación simplemente se procede a ejecutar de nuevo el algoritmo de inferencia. Para hacer frente a estos casos en los que las RDBs se hacen muy grandes rápidamente (sobre todo cuando la duración del segmento de tiempo es corto), se incorpora un proceso de razonamiento consistente en una "ventana" deslizante, de tamaño fijo que se mueve hacia adelante con el tiempo. Cuando se desplaza la "ventana" hacia delante, un corte de tiempo se sale y uno nuevo es añadido. Con este uso de "ventaja" fija, cada vez que se avanza en el tiempo las observaciones anteriormente recibidas no estarán directamente disponibles. En su lugar, se tendrá la creencia actual de los nodos, que pasará posteriormente a ser las siguientes probabilidades a priori.

La figura 3.5 muestra el proceso anteriormente explicado en una RDB que consta de una ventana con dos cortes de tiempo. Esta estructura puede ser considerada como un RBD genérica que consiste en el nodo estado X, su correspondiente nodo de observación O, donde la evidencia es añadida y un nodo acción A, que afectará al nodo de estado en el siguiente tiempo. Esto es muy importante pues es lo que se ha usado para el desarrollo de nuestro proyecto.

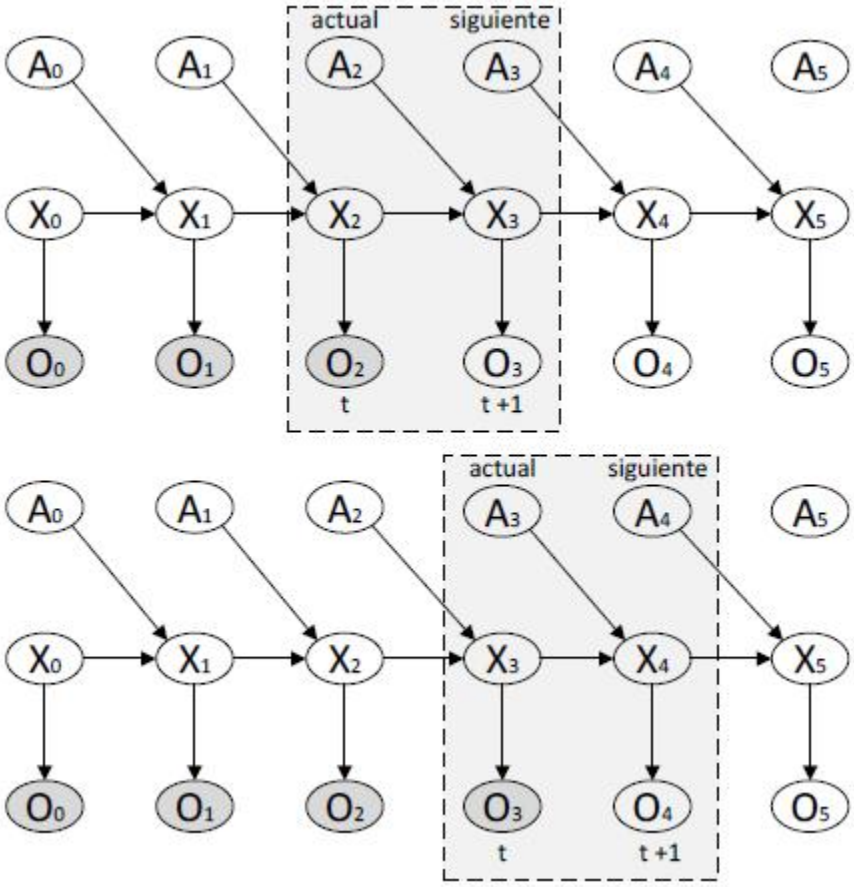

Figura 3.5: Ejemplo simplificación RDB

En resumen, si tenemos muchos cortes de tiempo (como es nuestro caso), apoyándonos en la suposición markoviana de que cada nodo sólo depende del anterior podemos simplificar las RDBs eligiendo cada vez dos estados y aplicando los métodos de inferencia a sólo esos dos. En cada paso, se irá deslizando la ventada para escoger los dos espacios temporales siguientes. En la figura 3.5 se puede apreciar este proceso.

3.4. Software

En el caso de este proyecto, se ha desarrollado un programa específico para la inferencia de la red dinámica de bayes con MatLab pero se podría haber escogido alguno de los programas que se pueden encontrar fácilmente por la red para realizar la inferencia. En este apartado se van a mencionar brevemente tres de los más importantes.

3.4.1. Elvira

Es un proyecto español, que surgió de la colaboración de cuatro universidades: Granada, Almeria, País Vasco y UNED y que se desarrollo principalmente entre 1997 y 2000. El objetivo principal del proyecto era la construcción de un entorno que entre otras cosas incluía la implementación de sistemas expertos bayesianos, el programa resultante se llamo Elvira.

El programa Elvira cuenta con un formato propio para la codificación de los modelos, un lector-intérprete para los modelos codificados, una interfaz gráfica para la construcción de redes con opciones específicas para modelos canónicos (puertas OR, AND, MAX, etc.), algoritmos exactos y aproximados (estocásticos) de razonamiento tanto para variables discretas como continuas, métodos de explicación del razonamiento, algoritmos de toma de decisiones, aprendizaje de modelos a partir de bases de datos, fusión de redes, etc. Elvira está escrito y compilado en Java, lo cual permite que pueda funcionar en diferentes plataformas y sistemas operativos (linux, MS-DOS/Windows, Solaris, etc.). La principal limitación del programa es que la búsqueda de resultados de investigación a corto plazo dificultó la aplicación de los principios de la metodología de desarrollo de software. El programa es de gran ayuda para comenzar con las redes de bayes pues es muy intuitivo y fácil de manejar. Más información en la página web del proyecto [10].

3.4.2. Nética

Netica es un entorno gráfico que contiene un conjunto de funciones para la descripción y la inferencia de redes bayesianas que se pueden utilizar para desarrollar aplicaciones compatibles con entornos de Microsoft Windows. Su interfaz es bastante intuitiva y amigable, además de que con ella es posible realizar una abducción total y/o parcial, mediante forma gráfica o textual. Se pueden realizar análisis de sensibilidades y permite en uso de variables continuas. Otra característica es que se permiten definir submodelos además de poder visualizar los resultados de los estados de cada nodo sobre el propio grafo de la red mediante diagramas de barras. Toda la información se puede encontrar en la página web del proyecto [9].

3.4.3. ToolBox para MatLab

BNT es una herramienta para MatLab desarrollada por Kevin Murphy entre los años 1997-2002, toda la información se puede encontrar en la página web [11]. En general, permite realizar casi cualquier tipo de construcción de red, inferencia y aprendizaje que se quiera, tanto de redes de bayes normales como dinámicas, también incluye resolución de modelos ocultos de Markov.

Capítulo 4

Metodología empleada

4.1. Introducción

En este capítulo se va a explicar toda la metodología empleada para la construcción de la red dinámica de bayes. El capítulo está formado por dos secciones. En la primera, se aclarará la construcción de la red explicando con detalle la ejecución del grafo y mencionando las probabilidades que se han de calcular (la obtención de estas se explicará mas detalladamente en el Apéndice 2). En la segunda sección, se hablará sobre la inferencia de la red. El etiquetado de vídeos se ha explicado en el Apéndice 1 para mayor comodidad. En todo el proceso se ha usado como herramienta el programa de cálculo MatLab.

4.2. Construcción de la red

Para explicar la construcción de la red dinámica de bayes, se van a seguir los pasos mencionados en el apartado 3.2.2.

4.2.1. Identificación de las variables del problema

Con identificar las variables del problema nos referimos a elegir los nodos del grafo, es decir, las variables aleatorias de nuestro problema. El primer nodo elegido será "Hay teléfono" pues el objetivo del proyecto, es llegar a decir si en el vídeo estudiado hay una persona hablando por teléfono o no. Los otros nodos has sido elegidos pensando en las acciones características que se pueden dar al hablar una persona por teléfono. Para mayor sencillez en el manejo de la información a cada nodo se le ha representado con una letra, a continuación se exponen los nodos elegidos:

- Nodo A: Persona sentada.
- Nodo B: Persona coge objeto.
- Nodo C: Mano a la altura del oído.
- Nodo D: Persona deja objeto.
- Nodo E: Persona en pie.
- Nodo F: Persona en movimiento.
- Nodo G: Persona tecleando
- Nodo H: Mano a la altura del bolsillo
- Nodo F: Hay teléfono.

Una vez escogidas las variables que forman el problema, se les deben asignar que valores toman, en este caso es sencillo. Para los nodos que representan acciones (A-H) podrán tomar dos valores: '1' si la acción aparece en el vídeo o '0' si la acción no aparece en el

vídeo. Para el nodo F, se tomará un '1' en el caso de que haya teléfono en el vídeo o un '0' si no hay teléfono en el vídeo.

4.2.2. Identificación de las relaciones

Para este apartado ha sido muy importante el siguiente ejemplo sacado de [10]. Para una mejor explicación de como lo hemos usado para nuestro proyecto se va a ir comparando este con nuestro caso. El enunciado del ejemplo es el siguiente:

"Suponga que es un guardia de seguridad en alguna instalación subterránea secreta. Quiere saber si está lloviendo hoy, pero su única información del mundo exterior se produce cada mañana cuando ve al director entrando con, o sin, paraguas."

En este caso, cada día que pasa, únicamente se tiene la evidencia de observación de si el director trae o no el paraguas siendo ésta nuestra variable de observación, y la variable de predicción, si está lloviendo o no. La adaptación a nuestro caso es que las evidencias son las variables aleatorias de la A a la H, es decir, si en el vídeo a estudio se produce la acción A o no se produce y así con el resto de acciones. Por otra parte, la variable de predicción será si hay teléfono o no (nodo F). La red bayesiana asociada al problema se presenta en la figura 4.1 y por analogía directa deducimos que la red bayesiana de nuestro caso será la mostrada en la figura 4.2.

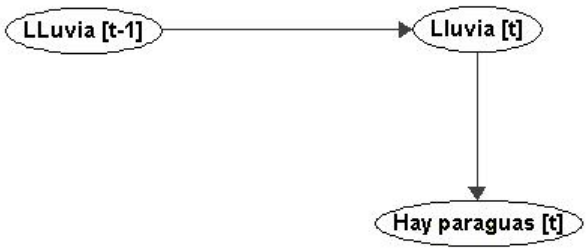

Figura 4.1: Red bayesiana de ejemplo

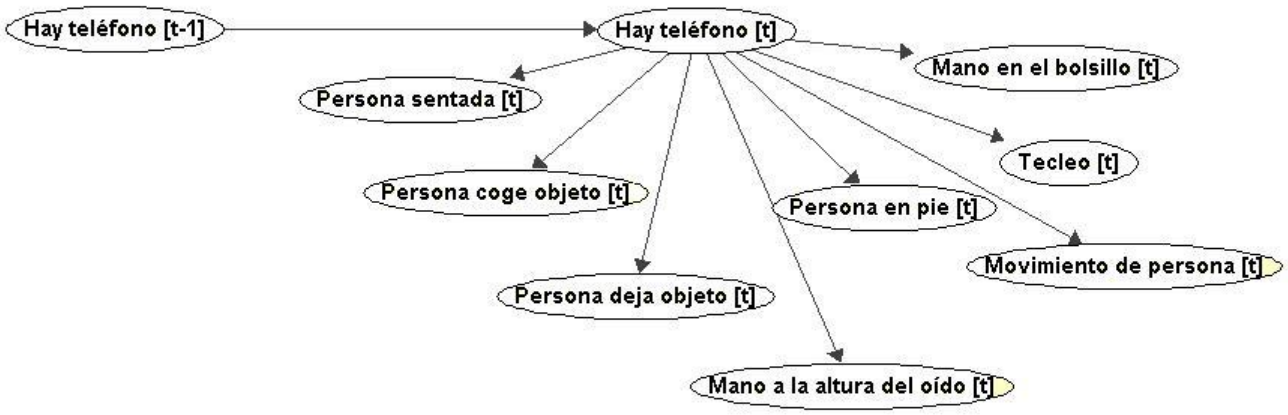

Figura 4.2: Red bayesiana de nuestro caso

4.2.3. Obtención de las probabilidades

Una vez que ya sabemos las relaciones entre las variables de nuestro problema podemos saber que probabilidades serán necesarias calcular.

- La probabilidad a priori de los nodos raíz. En nuestro caso el nodo raíz es "Hay teléfono" y la probabilidad que tenemos que calcular es la de que haya teléfono en la imagen.
- Las probabilidades condicionadas. Serán las probabilidades de que sabiendo que hay teléfono se den las evidencias.
- Las probabilidades temporales. Serán las probabilidades de que sabiendo que en t-1 ha ocurrido un suceso, ocurra otra vez en t ó que sabiendo que en t-1 se ha dado una variable, se de otra variable en t.

El cálculo de probabilidades se explicará con mas detalle en el apéndice B.

4.3. Inferencia de la red

Con la inferencia de la red lo que se quiere conseguir es la probabilidad de que haya teléfono en el vídeo una vez introducidas las evidencias.

Se denotará como la abreviación Tel(t) a la variable teléfono en un instante t de tiempo y Evidencias(t) a las evidencias en cada tiempo t. De la observación de la red se obtiene que la relación temporal será entre las variables Tel(t-1) y Tel(t) tal que p(Tel(t)|Tel(t-1)) (probabilidades temporales calculadas a partir de los vídeos) y el modelo de observación será entre las variable Tel(t) y Evidencias(t) de modo que p(Evidencias(t) |Tel(t)) (probabilidades condicionales calculadas a partir de los vídeos).

1. Para la primera diapositiva, t=1, el cálculo de la probabilidad será: p(Tel(1))=p(Tel(0)), siendo P(Tel(0)) la probabilidad a priori calculada a partir de los vídeos. Ahora introducimos las evidencias para t=1:

p(Tel(1)=si|Evidencias(1))=p(Evidencias(1)|Tel(1)=si)

y p(Tel(1)=no|Evidencias(1))=p(Evidencias(1)|Tel(1)=no).

Estas dos últimas probabilidades son las que se usará en t=2 pero antes es necesario normalizarlas pues la suma de ambas debe ser siempre 1.

2. Para la segunda diapositiva, t=2, sin introducir evidencias, la primera probabilidad a calcular será: p(Tel(2)|Evidencias(1))=p(Tel(1)|Evidencias(1)). Introducimos las evidencias en el t=2:

p(Tel(2)=si|Evidencias(1),Evidencias(2))=p(Evidencias(2)|Tel(2)=si)·p(Tel(2)=si|Evidencias(1))
p(Tel(2)=no|Evidencias(1),Evidencias(2))=p(Evidencias(2)|Tel(1)=no).·p(Tel(2)=no|Evidencias(1)).

Esto se repite para todas las diapositivas. Si la probabilidad final es p(Tel(t))=1, entonces habrá teléfono. Si por el contrario es 0 entonces no habrá teléfono. Es muy importante, al final de cada paso normalizar las probabilidades porque si no nunca convergerán a '1' o '0'.

Capítulo 5

Experimentación

Una vez construida la red toca comprobar si los resultados obtenidos son los esperados. En este caso, se usaron 33 vídeos, de los cuales 17 se utilizaron para construir los casos y calcular las probabilidades necesarias para las construcción de la red y los otros 16 se usaron como evidencias.

En este apartado no se expondrán todos los resultados obtenidos, solo los más importantes y representativos.

5.1. Vídeo 29

El vídeo 29, es un vídeo en el que aparece la siguiente secuencia de acciones: persona sentada, coge un teléfono, se lleva la mano a la altura del oído para hablar con el y luego deja el teléfono. Esta formado por 525 diapositivas, las probabilidades temporales normalizadas que nos da MatLab tras la inferencia se representan en el gráfico de la figura 5.1. En azul la probabilidad representada es la de que haya teléfono en el vídeo y en verde la contraria. Como se puede observar para la diapositiva número 100 ya es claro

que en el vídeo hay teléfono.

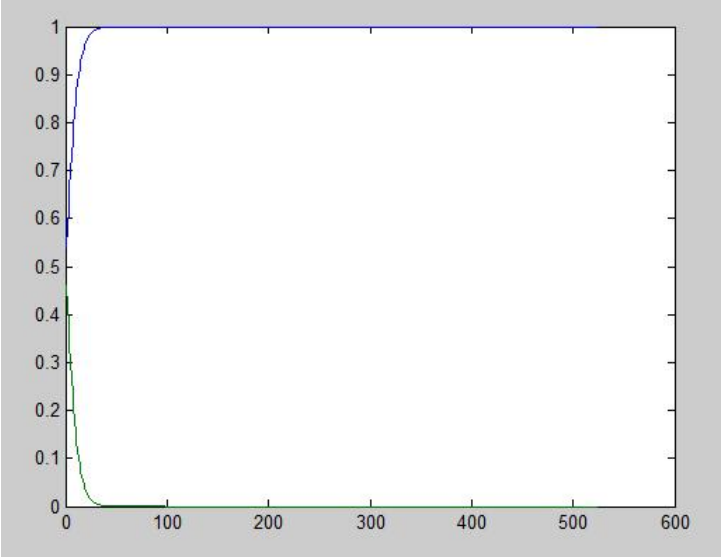

Figura 5.1: Probabilidades temporales del vídeo 29

5.2. Vídeo 132

El vídeo 132, es un vídeo en el que aparece la siguiente secuencia de acciones: persona sentada, se lleva la mano al bolsillo para sacar el móvil, teclea algo y se lleva la mano a la altura del oído para hablar con el. Esta formado por 623 diapositivas, las probabilidades temporales normalizadas que nos da MatLab tras la inferencia se representan en el gráfico de la figura 5.2. En azul, la probabilidad representada es la de que haya teléfono en el vídeo y en verde la contraria. Como se puede observar ya para antes de la diapositiva número 100 es claro que en el vídeo hay teléfono. Se destaca que aunque la secuencia de acciones es diferente al caso anterior, el método sigue funcionando.

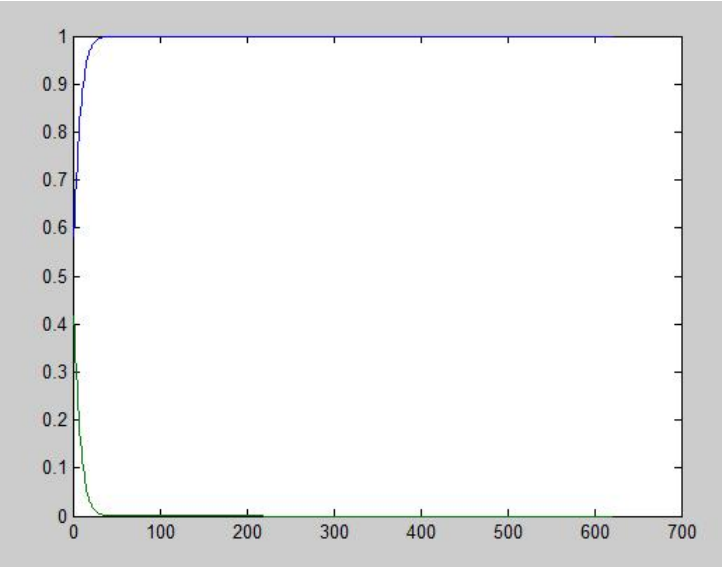

Figura 5.2: Probabilidades temporales del vídeo 132

5.3. Vídeo 62

El vídeo 62, es un vídeo en el que aparece la siguiente secuencia de acciones: persona de pie en movimiento, que en en un determinado momento coge algo que no es un móvil. Esta formado por 207 diapositivas, las probabilidades temporales normalizadas que nos da MatLab tras la inferencia se representan en el gráfico de la figura 5.3. En azul, la probabilidad representada es la de que haya teléfono en el vídeo y en verde la contraria. Como se puede observar con muy pocas diapositivas ya indica que no hay teléfono en el vídeo, para explicar mejor esto es conveniente fijarnos en la figura 5.4 que representa las probabilidades temporales sin normalizar. Como se puede observar la probabilidad de que haya teléfono siempre está rondando el cero y la probabilidad de que no haya teléfono no es muy alta pero con la normalización se convierte rápidamente en 1.

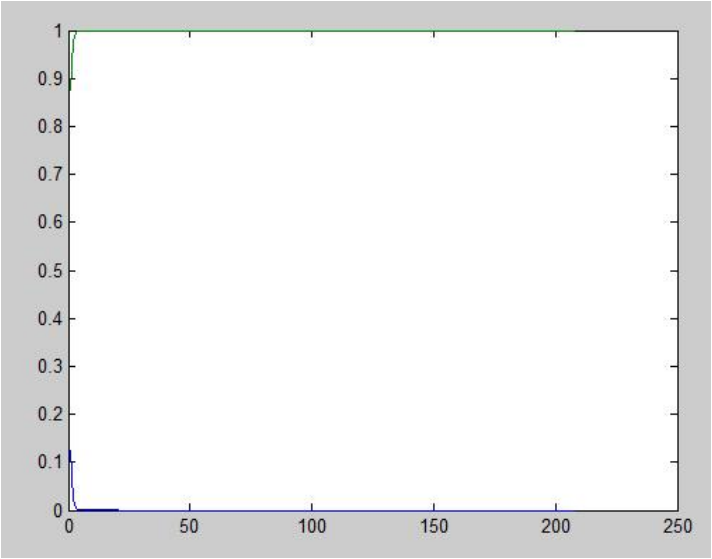

Figura 5.3: Probabilidades temporales del vídeo 62

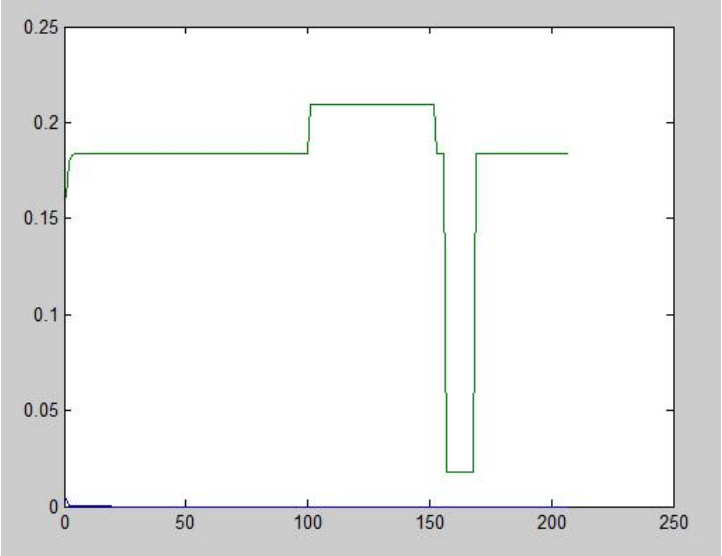

Figura 5.4: Probabilidades temporales sin normalizar del vídeo 62

5.4. Vídeo 25

El vídeo 25, es un vídeo en el que aparece la siguiente secuencia de acciones: persona de pie, coge un teléfono de la mesa, se lo lleva al oído para hablar y luego lo deja sobre

la mesa. Esta formado por 300 diapositivas, las probabilidades temporales normalizadas que nos da MatLab tras la inferencia se representan en el gráfico de la figura 5.5. En azul, la probabilidad representada es la de que haya teléfono en el vídeo y en verde la contraria. Este gráfico es un poco diferente pues primero la probabilidad de que haya teléfono es cero pero a la altura de la diapositiva 50 cambia radicalmente a uno. Decir que este cambio no es tan radical pero con la normalización se hace así.

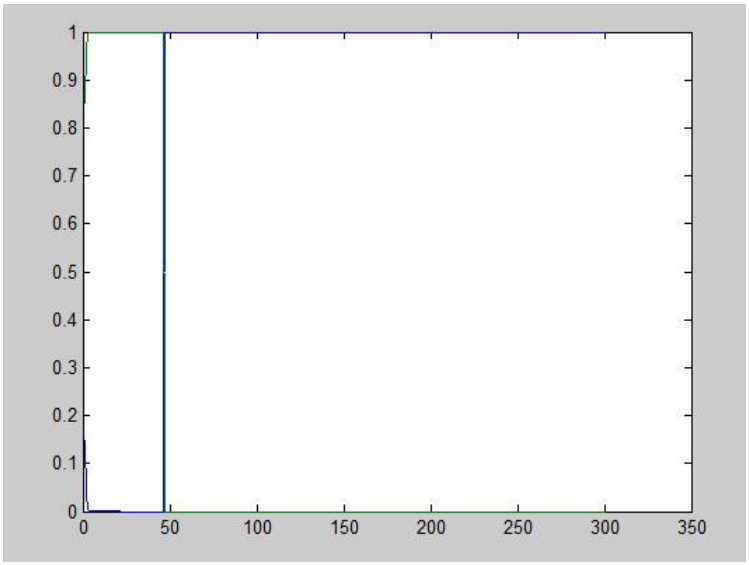

Figura 5.5: Probabilidades temporales del vídeo 25

5.5. Vídeo 26

El vídeo 26, es un vídeo en el que aparece la siguiente secuencia de acciones: persona sentada, coge un teléfono de la mesa, se lo lleva al oído para hablar y luego lo deja sobre la mesa. Esta formado por 343 diapositivas, las probabilidades temporales normalizadas que nos da MatLab tras la inferencia se representan en el gráfico de la figura 5.6. En azul, la probabilidad representada es la de que haya teléfono en el vídeo y en verde la contraria. Se observa que sobre la diapositiva 50 ya converge a un valor válido de que si

hay teléfono en el vídeo.

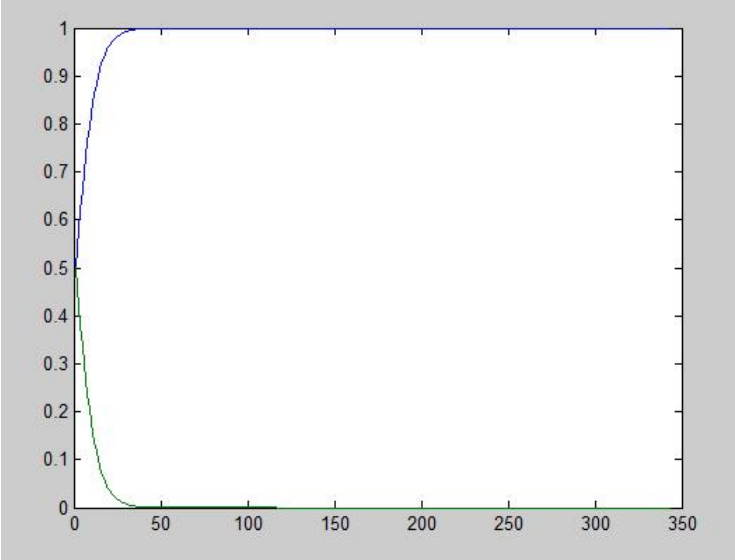

Figura 5.6: Probabilidades temporales del vídeo 26

5.6. Conclusión

En general se puede observar que en seguida se converge a un resultado válido. Todos los vídeos usados como evidencia funcionan, dando el resultado correcto de si hay teléfono o no.

Capítulo 6

Conclusiones

Finalizando este proyecto, decir que se puede ampliar en un futuro para reconocer cualquier tipo de objeto que tenga un uso característico, como podrían ser libros, lápices o mochilas. Lo único diferente serían los nodos, es decir, las acciones que distinguen el uso del objeto, por lo demás, el grafo sería el mismo y las probabilidades habría que calcularlas a partir de un nuevo etiquetado de vídeos.

También se puede ampliar el proyecto de forma que el etiquetado de vídeos sea automático y no manual. Como se hablo en el apartado 2.3.2, la idea sería crear otras dos redes dinámicas de bayes, una para el reconocimiento de las acciones y otra para la estimación de la pose humana.

Por otra parte, destacar que la red dinámica de bayes funciona bastante bien, pues en la mayoría de los casos, a partir de la diapositiva 100 ya se sabe si hay teléfono en el vídeo o no.

Bibliografía

[1] Javier García de Jalón, José Ignacio Rodríguez, Jesús Vidal. *Aprenda MatLab 7.0 como si estuviera en primero.* Madrid: Diciembre, 2005.

[2] Francisco J. Ruiz-Ruano Campaña. *LyX: Con "L" de LaTeX.* Octubre, 2009.

[3] M. Valera and S.A. Velastin. *Intelligent Distributed Surveillance Systems*, IEEE Proc.-Vis. Image Signal Processing., vol. 152, no. 2, Apr. 2005.

[4] Hiroshi Miki, Atsuhiro Kojima, and Koichi Kise. Environment Recognition Based on Human Actions Using Probability Networks.*International Journal of Smart Home* .Vol.3, No.3, July 2009

[5] C. Bregler. Learning and recognizing human dynamics in video sequences. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* 1997

[6] S. Fine, Y. Singer, and N. Tishby. The hierarchical hidden markov model: Analysis and applications. *Machine Learning*, 32(1):41-62, 1998.

[7] Russell, Stuart y Norvig, Peter. *Inteligencia Artificial.Un enfoque moderno.* Madrid:Pearson Prentice Hall, 2004.

[8] Gary Bradsky, Adrian Kaebler. *Learning OpenCV. Computer Vision with the OpenCV Library.*

[9] http://www.norsys.com/

[10] http://leo.ugr.es/elvira/

[11] https://code.google.com/p/bnt/

Parte II

Apéndices

Apéndice A

Etiquetado de vídeo

En este anexo se explicará con detalle como se han etiquetado los 110 vídeos que se han usado en este proyecto. Se deben explicar ciertos aspectos para entender bien el siguiente apartado:

- En algunos vídeos aparece más de una persona, pero el etiquetado se ha hecho sólo respecto a una, en general, la que se ha encontrado más interesante para nuestro proyecto. Por ejemplo, si en un vídeo aparece una persona hablando por teléfono y en segundo plano una persona leyendo un libro, sólo se habrá etiquetado la persona que interactúa con el teléfono.

- No se han etiquetado todos los vídeos pues en muchos de ellos, la secuencia de acciones se repite y no constituyen un aporte de información a nuestro proyecto, por tanto, se ha desestimado el etiquetado de estos.

- Hay que diferenciar que no todos los vídeos se han usado para lo mismo. Hay un grupo de vídeos que se han usado para construir los diferentes casos del proyecto y calcular las probabilidades condicionales y temporales con ellos. Mientras que el

resto de vídeos se han usado como prueba para ver si el proyecto funcionaba.

El etiquetado de los vídeos se ha hecho con MatLab, teniendo cada vídeo su propio fichero. De esta forma, cada uno de estos construye una matriz con un número de filas igual al número de diapositivas, y un número de columnas igual a las variables aleatorias de nuestro problema (9 nodos). Esta matriz esta compuesta sólo de ceros y unos, de tal manera que si un '1' aparece en la casilla indica que la acción de la columna se está dando en la diapositiva de la fila y si aparece un '0' será que la acción no está ocurriendo. Así se consigue saber fácilmente las acciones que están ocurriendo en cada momento en los vídeos.

El reconocimiento de estas acciones ha sido hecho de forma visual, a continuación se muestra que tipo de imágenes se han asociado a los diferentes nodos.

Figura A.1: Nodo A: Persona sentada

Figura A.2: Nodo B: Coger Objeto

Figura A.3: Nodo C: Mano altura del oído

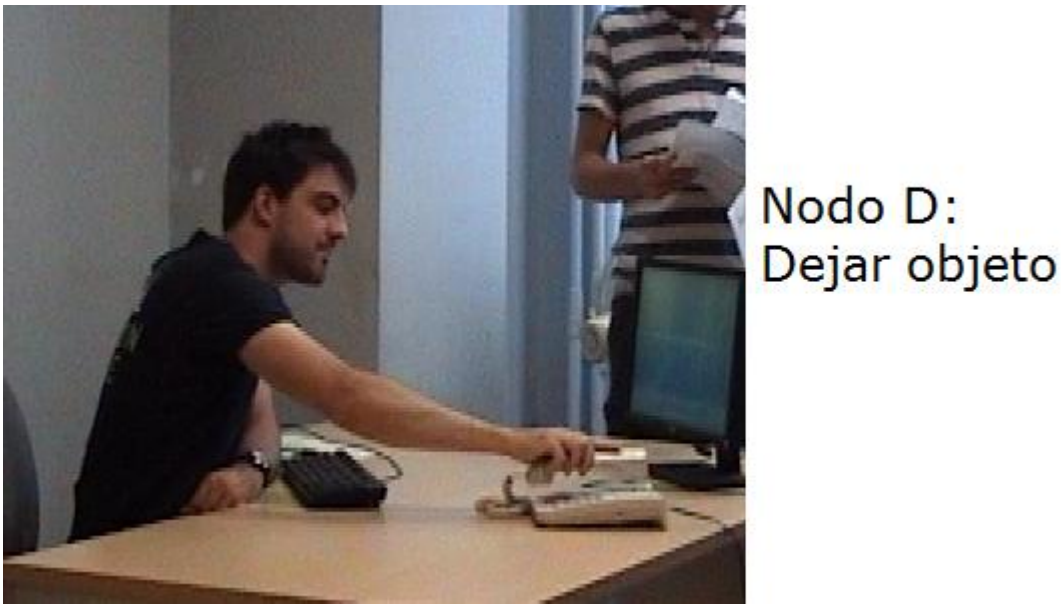

Figura A.4: Nodo D: Dejar objeto

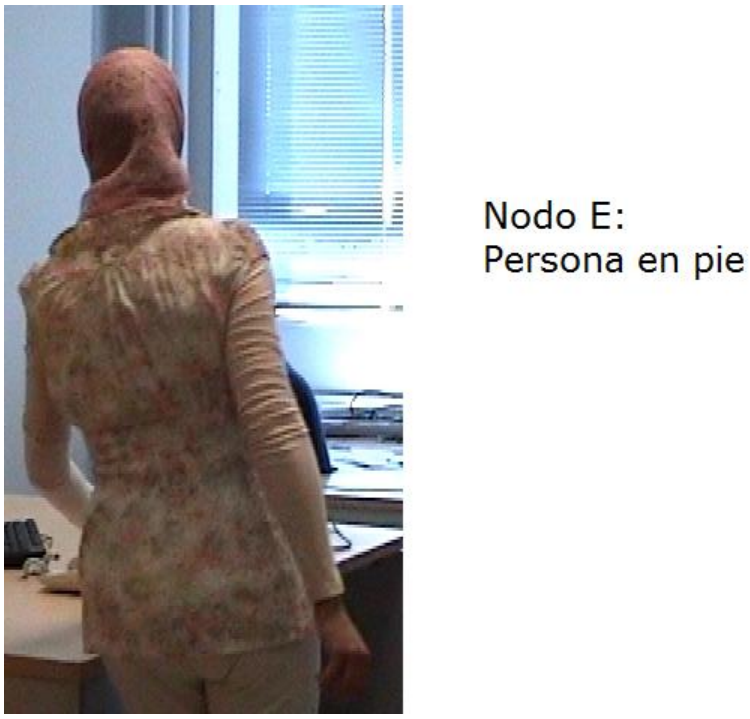

Figura A.5: Nodo E: Persona en pie

Figura A.6: Nodo F: Persona en movimiento

Figura A.7: Nodo G: Teclea

Figura A.8: Nodo H: Mano altura del bolsillo

En la siguiente tabla se han clasificado los vídeos por número, objeto que aparece en él y uso que se le ha dado en el proyecto.

Nº de Vídeo	Objeto	Uso	Descripción uso/no uso
1	Mochila	No	Sólo aparecen personas moviéndose
3	Libro	Si	Para casos
4	Libro	Si	Para prueba
8	Libro	No	No aporta información
9	Puerta	No	Sólo vamos a tener en cuenta el hombre de la mochila
9	Mochila	SI	Para casos
11	Mochila	No	No aporta información
12	Libro	Si	Para casos

Nº de Vídeo	Objeto	Uso	Descripción uso/no uso
14	Libro	Si	Para prueba
15	Puerta	Si	Para prueba
16	Puerta	No	Redundante
17	Puerta	Si	Para prueba
19	Mochila	Si	Para prueba
20	Libro	No	Sólo se etiqueta la persona que usa el teléfono
20	Ordenador	No	Sólo se etiqueta la persona que usa el teléfono
20	Teléfono	Si	Para caso
22	Libro	No	Sólo se etiqueta la persona que usa el teléfono
22	Ordenador	No	Sólo se etiqueta la persona que usa el teléfono
22	Teléfono	Si	Para caso
23	Libro	No	Sólo se etiqueta la persona que usa el teléfono
23	Ordenador	No	Sólo se etiqueta la persona que usa el teléfono
23	Teléfono	Si	Para caso
25	Teléfono	Si	Para prueba
26	Ordenador	No	Redundante
26	Teléfono	Si	Para prueba
28	Ordenador	No	Sólo se etiqueta la persona que usa el teléfono
28	Teléfono	Si	Para caso
29	Ordenador	No	Sólo se etiqueta la persona que usa el teléfono

Nº de Vídeo	Objeto	Uso	Descripción uso/no uso
29	Teléfono	Si	Para prueba
33	Pizarra	No	Redundante
34	Pizarra	Si	Como prueba
36	Libro	No	Redundante
37	Libro	No	Redundante
38	Libro	No	Redundante
40	Ordenador	Si	Como caso
42	Ordenador	No	Sólo se etiqueta la persona que usa el teléfono
42	Teléfono	Si	Para casos
44	Libro	No	Redundante
45	Libro	No	Redundante
46	Libro	No	Redundante
48	Puerta	No	Redundante
49	Puerta	No	Redundante
51	Puerta	No	Redundante
52	Puerta	No	Redundante
53	Puerta	si	Para prueba
55	Puerta	No	Redundante
56	Puerta	No	Redundante
58	Puerta	No	Redundante

Nº de Vídeo	Objeto	Uso	Descripción uso/no uso
59	Puerta	No	Redundante
61	Mochila	Si	Para caso
62	Mochila	Si	Para prueba
64	Mochila	No	Redundante
68	Puerta	Si	Para prueba
69	Puerta	No	Redundante
70	Puerta	No	Redundante
72	Puerta	No	Redundante
73	Puerta	No	Redundante
74	Puerta	No	Redundante
76	Puerta	No	Redundante
77	Puerta	No	Redundante
79	Pizarra	No	Redundante
80	Pizarra	No	Redundante
81	Ordenador	Si	Para casos
83	Pizarra	No	Redundante
84	Ordenador	No	Redundante
86	Pizarra	Si	Para casos
87	Ordenador	No	No hay personas en escena
87	Ordenador	No	No hay personas en escena

Nº de Vídeo	Objeto	Uso	Descripción uso/no uso
88	Puerta	Si	Para casos
89	Puerta	No	Redundante
90	Puerta	No	Redundante
91	Puerta	No	Redundante
92	Puerta	No	Redundante
94	Puerta	No	Redundante
101	Puerta	No	Redundante
102	Puerta	No	Redundante
103	Puerta	No	Redundante
107	Puerta	No	Redundante
108	Puerta	No	Redundante
110	Puerta	No	Redundante
113	Puerta	No	No aporta información
115	Puerta	No	No aporta información
116	Puerta	No	No aporta información
118	Pizarra	No	Redundante
118	Puerta	No	No aporta información
118	Mochila	Si	Como prueba
119	Pizarra	No	No aporta información
119	Mochila	No	Redundante

Nº de Vídeo	Objeto	Uso	Descripción uso/no uso
119	Puerta	No	Redundante
121	Mochila	No	Redundante
121	Pizarra	No	Sólo se etiqueta la persona que usa el teléfono
121	Teléfono	Si	Como caso
122	Mochila	Si	Como caso
124	Mochila	No	Sólo se etiqueta la persona que usa el teléfono
124	Pizarra	No	Sólo se etiqueta la persona que usa el teléfono
124	Teléfono	Si	Como prueba
125	Pizarra	No	Redundante
127	Pizarra	No	No aporta información
128	Pizarra	No	No aporta información
128	Puerta	No	Redundante
129	Pizarra	No	Sólo se etiqueta la persona que usa el teléfono
129	Puerta	No	Sólo se etiqueta la persona que usa el teléfono
129	Teléfono	Si	Como prueba
131	Ordenador	No	Redundante
132	Ordenador	No	Redundante
132	Teléfono	Si	Como prueba
134	Ordenador	Si	Como prueba
135	Mochila	No	Redundante

Nº de Vídeo	Objeto	Uso	Descripción uso/no uso
135	Ordenador	Si	Como caso
136	Libro	No	Redundante
138	Libro	No	Redundante
139	Mochila	No	No aporta información
139	Mochila	No	No aporta información
140	Mochila	No	No aporta información
140	Libro	No	No aporta información
142	Libro	No	No aporta información
144	Libro	No	No aporta información
148	Mochila	No	No aporta información
149	Mochila	No	No aporta información
152	Mochila	No	No aporta información
152	Teléfono	No	No aporta información
154	Puerta	No	Redundante
155	Puerta	No	Redundante
157	Puerta	No	Redundante
159	Puerta	No	Redundante
160	Puerta	No	Redundante
162	Puerta	No	Redundante
163	Puerta	No	Redundante

Nº de Vídeo	Objeto	Uso	Descripción uso/no uso
164	Libro	No	Redundante
166	Puerta	No	Redundante
167	Libro	No	Redundante
167	Puerta	No	Redundante

Apéndice B

Obtención de probabilidades

- **La probabilidad a priori de los nodos raíz.** Estas probabilidades son las más sencillas de calcular, son sólo dos, serán:

$$p(Tel = si) = \frac{n^{\underline{o}} \text{ de diapositivas en las que hay teléfono}}{n^{\underline{o}} \text{ de diapositivas totales}}$$

$$p(Tel = no) = 1 - p(Tel = si)$$

.

Los resultados obtenidos han sido: p(Tel=si)=0.457 y p(Tel=no)=0.5433.

- **Las probabilidades condicionadas.** Estas probabilidades son p(Evidencias(t)|Tel(t)=si) y p(Evidencias(t)|Tel(t)=no). A cada combinación de evidencias posibles vamos a llamarlo "caso" y tendremos que calcular las probabilidades de que se den esos casos tanto si hay teléfono como no. Como el número de evidencias es 8 (una por cada nodo), se llega a la conclusión de que el número combinaciones que tendríamos

sería 2^8 para tel=si y otras tantas para tel=no. Calcular todas estas probabilidades es difícil, pero hay que darse cuenta de que realmente no son tantos casos, pues hay muchos nodos que no pueden darse a la vez, es decir, no tienen relación. Por ejemplo: mano a la altura del oído y mano en el bolsillo no se pueden dar a la vez, y ocurre lo mismo con los nodos persona sentada y persona de pie. Eliminando esos casos, llegamos a la conclusión de que sólo hay 19 combinaciones posibles de las que tendremos que calcular probabilidades. En la tabla adjunta se representa el número de casos con los nodos que forman cada uno y las probabilidades que se han obtenido.

Nº de Caso	A	B	C	D	E	F	G	H	Probabilidad con tel=si	Probabilidad con tel=no
1	1	1	0	0	0	0	0	0	0,0206	0
2	1	0	1	0	0	0	0	0	0,279	0
3	1	0	0	1	0	0	0	0	0,024	0
4	1	0	0	0	0	0	1	0	0,0275	0
5	1	0	0	0	0	0	0	1	0,012	0
6	1	0	0	0	0	0	0	0	0,3964	0,335
7	0	1	0	0	1	1	0	0	0,0058	0
8	0	1	0	0	1	0	0	0	0,0017	0,0147
9	0	0	1	0	1	1	0	0	0,0199	0
10	0	0	1	0	1	0	0	0	0,0959	0
11	0	0	0	1	1	1	0	0	0	0,0182
12	0	0	0	1	1	0	0	0	0,012	0,0523

Nº de Caso	A	B	C	D	E	F	G	H	Probabilidad con tel=si	Probabilidad con tel=no
13	0	0	0	0	1	1	1	0	0	0
14	0	0	0	0	1	1	0	1	0,0099	0,0161
15	0	0	0	0	1	1	0	0	0,0306	0,1835
16	0	0	0	0	1	0	1	0	0	0
17	0	0	0	0	1	0	0	1	0,0137	0,0196
17	0	0	0	0	1	0	0	0	0,0354	0,2093
19	0	0	0	0	0	0	0	0	0,0134	0,065

Para el cálculo de las probabilidades simplemente se ha usado esta fórmula:

$$p(Caso) = \frac{n^{\underline{o}} \text{ de diapositivas en las que aparece el caso}}{n^{\underline{o}} \text{ de diapositivas totales}}$$

- **Las probabilidades temporales.** Estas son diferentes para cada vídeo que introducimos como evidencia, se calculan en cada iteración y se tendrán en cuenta que la primera probabilidad temporal será la probabilidad a priori y luego se irán propagando de forma que p(Tel(t))=p(Tel(t-1)). Habrá que calcular tantas como diapositivas tenga el vídeo y la última será la que nos indique si hay teléfono en el vídeo o no.

- **Normalización de las probabilidades**. Este proceso es importante porque si no, nunca se llegaría a tener como resultado un '1' o un '0', el cálculo es simple, sólo

hay que tener en cuenta que p(Tel(t)=si)+p(Tel(t)=no)=1.

$$p(Tel(t) = si) = \frac{p(Tel(t) = si)}{p(Tel(t) = si) + p(Tel(t) = no)}$$

$$p(Tel(t) = no) = \frac{p(Tel(t) = no)}{p(Tel(t) = si) + p(Tel(t) = no)}$$